Finding Fiction: Facilitating Access to Works of the Imagination Scattered by Form and Format

Abstract: This study explores ways to assist users who are primarily interested in finding a good story, regardless of format or literary form. The emphasis is on materials classed in the Arts (700s) and Literature (800s) in the Dewey Decimal Classification (DDC) system. Features from two prototypes, FictionFinder and DeweyBrowser, are being combined to attempt to assist users in finding terms to input while providing a holistic approach to finding works with imaginary content.

1. Introduction

Wikipedia gives both broad and narrow definitions of “fiction”:

Fiction is storytelling of imagined events and stands in contrast to non-fiction, which makes factual claims about reality. . . . Fictional works—novels, stories, fairy tales, fables, films, comics, interactive fiction—may be partly based on factual occurrences but always contain some imaginary content. The term is also often used synonymously with literature and more specifically fictional prose. In this sense, fiction refers only to novels or short stories . . . . (Fiction, 2006).

In DDC 22 (Dewey, 2003), “fiction” has the narrow definition, limited to prose. The table of preference in the schedule at 800 and the discussion of literary form in the Manual at 800 make clear that stories presented in the form of drama or poetry are classed with drama or poetry, not with fiction. The DDC is structured so that purely textual literary forms are classed with literature in the 800s, but formats that combine literary text with other arts are classed with the arts in the 700s, e.g., comic books, opera, theater, and films (Beall, 2004). Format is given precedence over content for fictional works in the arts and literature. Factual works are classed with the topic, e.g., films that teach history in the 900s. Comic books and graphic novels are examples of a mixed format classed with the arts, specifically under drawing and drawings in 741.5 Comic books, graphic novels, fotonovelas, cartoons, caricatures, comic strips (Dewey Updates, 2006). Goldsmith defines “graphic novel” as “a part of a spectrum of sequential art formats that includes a range of related media such as one-panel gag cartoons and serially published comic books.” She focuses on “the graphic novel element of the spectrum, …on creative works that include narrative with a beginning, a middle, and an end and are published in book format. . . .” (Goldsmith, 2005, 16). Discussions about how to improve provision for graphic novels in the DDC have revealed strong differences of opinion about whether graphic novels belong with the arts in the 700s or with literature in the 800s; some librarians argue that readers of graphic novels are primarily interested in the art; others that they are primarily interested in narrative fiction (Beall, 2004). Goldsmith lists four common ways that fictional graphic novels are handled in libraries; two involve putting them outside the general classification scheme; one is placement in 741.5; and the other is to place them “in the fiction, or science fiction, or short story collection—either in the literature classification or, as numerous libraries maintain their fiction collections, according to the author’s surname” (Goldsmith, 2005, 53-54).
2. Finding Fiction
Most general libraries will have some users especially interested in the artistic aspect of mixed formats, and others in the literary aspect. Consequently, local solutions by which some libraries place mixed formats in the 700s and some place them in the 800s will never satisfy all users; and users who visit multiple libraries are likely to be confused. This study explores ways to assist users who are primarily interested in finding a good story, regardless of format or literary form. The emphasis is on the 700s and 800s. In the 700s, the focus is on comic books and graphic novels (741.5), films (791.43), and television programs (791.45). Other fictional formats in the 700s (e.g., operas in 782.1) are excluded from this project for now, though ultimately they will need to be included. Similarly, folk tales, classed with folklore in 398.2, are excluded from this project for now unless they are identified as fiction by an accepted value for literary form (see section 4). Our approach is to consider ways to combine key features of projects like DeweyBrowser and FictionFinder.

3. Fiction Records in WorldCat
This project seeks to build on much previous work, first and foremost the cataloging records contributed to WorldCat by catalogers from around the world. These records include author and title; publisher and publication date; format (e.g., audiobook, large print); language; translator; summaries or other annotations; links to outside information, e.g., authors' web sites; cover art; library holdings, pseudonyms and uniform titles from authority records; DDC or Library of Congress Classification (LCC) class number or both; subject headings for themes, settings, characters; genre headings. The WorldCat database reflects the extra efforts that catalogers have been making to improve subject and genre access to fiction since publication of *Guidelines on Subject Access to Individual Works of Fiction, Drama, Etc.* (GSAFD; American Library Association, first edition 1990; second edition 2000). For example:

Since 1997 the British Library has adopted a policy for providing enhanced subject access to individual works of fiction catalogued for the British National Bibliography. Genre headings are applied where appropriate in accordance with the *Guidelines on Subject Access to Individual Works of Fiction, Drama, Etc.*” (British Library, 2006).

In the OCLC WorldCat database, more than 500,000 genre terms based on GSAFD or Library of Congress Subject Headings have been assigned by catalogers.

The amount of data beyond the basic minimum varies from record to record. For the FictionFinder prototype, the OCLC FRBR Work-Set algorithm has been applied to cluster the records at the work level. Records with the same author/title key are members of the same work cluster (Hickey, O’Neill & Toves, 2002). In the FictionFinder interface the term “edition” is used instead of the FRBR “manifestation” because “edition” will be more readily understood by library users. We follow that usage in this paper. Because of FRBRization, many edition records with minimal data are linked to work records rich with data gathered from all the edition records; as a result, the edition records can be found via the rich data in the work records. Some efforts have been made to add data algorithmically, e.g., to supply DDC numbers for work records lacking them. Because of the large number of records involved (2.8 million edition records and 1.4 million works), this project cannot contemplate adding data manually.

This project utilizes all the subject data available in WorldCat except the LCC class numbers. The LCC numbers could be utilized in a later version, for example, to provide more information about the nationality of authors, e.g., distinguishing Colombian, Mexican, Spanish authors.
Clare Beghtol, Annelise Mark Pejtersen, and others have discussed the value of providing much more information about fictional works, in the form of a classification scheme or other controlled vocabulary. For example, Beghtol proposed giving the following information for all major characters: name; occupation; religion; socioeconomic group; racial, ethnic, national group; physical or mental health; sex; age; living or not; type (e.g., human beings) (Beghtol, 1994, 253-266). For another example, Pejtersen proposed describing an author's intention to convey information or to create a particular kind of emotional experience; this information would be presented in standard phrases available for searching (Pejtersen, 1994, 258-261). One sample book had in the “information” category “A description of allegiance and patriotism” and in the “experience” category “Exciting” (Rasmussen, 1994, 300). Unfortunately, the detailed representation of fictional works envisioned by Beghtol, Pejtersen, and others is not present in the WorldCat database, at least not in any systematic, controlled fashion. The current project cannot draw any conclusions about the potential value of data not present in the WorldCat database.

One advantage to using the WorldCat database, in addition to the large number of bibliographic records, is the large amount of library holdings information (over 1 billion library holdings symbols) and the ability to build on the OCLC “Find-in-a-Library” service. Users who identify an edition of a work that interests them can easily find out which local libraries have the edition and search for it in the local library catalog (or identify libraries holding the edition in the same state, province, region, or worldwide and search the relevant catalogs). The holding library need not use Dewey or any other classification scheme for this feature to work.

4. DeweyBrowser and FictionFinder
4.1 DeweyBrowser

DeweyBrowser is a visual interface that displays search results in successive rows of ten categories based on the three main summaries of the DDC. A user enters a search and navigates up and down the Dewey hierarchy by clicking on a category. Categories are color-coded to indicate where matching records occur. Three collections are currently available for browsing using the DeweyBrowser, including wcat, a database of 2.2 million records for the most widely held works in WorldCat (Vizine-Goetz & Hickey, 2006).

DeweyBrowser has the advantage of presenting results of searches in a graphic display that helps users learn where in the DDC the good stories are found. DeweyBrowser can do that now for users who know what terms to input. For example, a search on the title words “pride prejudice” yields 9 records under 823 English fiction for Jane Austen's novel *Pride and Prejudice* itself, plus other records in the 800s for criticism, sequels, and dramatic adaptations. Under 791 Public performances the search yields 3 records for film adaptations (3 different films, including one “modern makeover” set in Utah) and 7 records for 2 BBC television adaptations. The 9 records for the novel and some of the records for the film and television adaptations can also be retrieved by the following topical subject headings:

- Courtship
- Sisters
- Young women

The 9 records for the novel and some of the records for the adaptations would also be retrievable by the following headings if genre and geographic subject headings were indexed in DeweyBrowser:
Domestic fiction
Love stories
England

Some of the records for the novel but not all can be retrieved by a search for the topical heading

Social classes

The DeweyBrowser does not aggregate records at the work level; if it did, 8 of the 9 editions of the novel itself could be retrieved by the same subject headings.

4.2 FictionFinder

FictionFinder employs a different approach. The database contains records for materials identified as fiction, novels, short stories, dramas, or comic strips and records for sound recordings for literary texts. Coding in the bibliographic records, MARC 21 field 008, is used to identify the specific literary forms included in the database (Library of Congress, 2004). The system presents records at the work level and aggregates information about fiction, such as names of characters, settings, genres, and subjects to assist users in finding works of fiction. For example, FictionFinder brings together more than 700 editions of Austen’s novel *Pride and Prejudice*, including illustrated and abridged editions and translations into more than 30 languages, under one work record with the key:

```
austen, jane\1775 1817/pride and prejudice
```

This work record and all the linked edition records can be found under all the headings mentioned above, plus other headings, for example:

Regency fiction
Bennet, Elizabeth (Fictitious character)
Darcy, Fitzwilliam (Fictitious character)
Marriage — 19th century

Because of an inconsistency in the title (“&” instead of “and”), 1 of the 9 records for the novel retrieved by DeweyBrowser is attached to a different work record, which brings together fewer than 15 editions under the key:

```
austen, jane\1775 1817/pride & prejudice
```

Films and television programs are not included in the existing FictionFinder prototype. Adding dramatic films on the basis of DDC numbers 791.43, 791.4334, 791.4372, 791.4375 (more than 9,000 work records) and dramatic television programs on the basis of 791.45, 791.4572, 791.4575 (more than 3,000 work records) will significantly increase the range of material to which FictionFinder provides access.

Comic books and graphic novels are included in FictionFinder only if they are specifically coded as fiction or comic strips. This coding is inconsistent for comic books and graphic novels, especially for older editions. A search in DeweyBrowser for “spider man” yields more than 60 records under 741 Drawing & drawings for works with DDC numbers 741.5+. Among those records are many for comic books republished as graphic novels that are not currently included in FictionFinder because they are not coded as fiction.
If there were no records for comic books and graphic novels already in FictionFinder, adding records on the basis of the DDC numbers 741.5, 741.58, 741.59+ would lead to an increase of more than 4,000 work records; however, since many of the work records are already in FictionFinder, the increase may be more significant in the number of editions and holdings to which FictionFinder provides access.

4.3 DeweyBrowser plus FictionFinder

By combining features from FictionFinder and DeweyBrowser we will attempt to assist users in finding terms to input while providing a holistic approach to finding works with imaginary content. Expanded criteria for selecting works of fiction, outlined in section 2, will be used to define the collection. Records for all types of materials will be derived from the OCLC WorldCat database and deployed through a modified DeweyBrowser interface. An early prototype that uses selected Dewey categories to narrow search results has been developed. An example of the results retrieved for a search for the subject heading “Courtship” is shown in Figure 1.

<table>
<thead>
<tr>
<th>Dewey Categories</th>
<th>Works from all Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ American fiction in English</td>
<td>The American / Henry James</td>
</tr>
<tr>
<td>□ English drama</td>
<td>Emma / Jane Austen</td>
</tr>
<tr>
<td>□ English fiction</td>
<td>High society / Sol C. Siegel; John Patrick; Charles Walters; Bing Crosby; Grace, Princess of Monaco; Frank Sinatra; Celeste Holm; John Lund; Louis Calhern; Sidney Blackmer; Louis Armstrong; Paul C. Vogel; Ralph E Winters; Cole Porter; Johnny Green; Saul Chaplin; Philip Barry; Metro-Goldwyn-Mayer; Warner Home Video (Firm)</td>
</tr>
<tr>
<td>□ Movies and films</td>
<td>Jane Austen's Pride and prejudice / Colin Firth; Jennifer Ehle; Alison Steadman; Benjamin Whitrow; Simon Langton; Andrew Davies; Sue Birtwistle; Michael Wearing; Jane Austen; Arts and Entertainment Network,; BBC Lionheart Television</td>
</tr>
<tr>
<td>□ Spanish fiction</td>
<td>Jude the obscure / Thomas Hardy</td>
</tr>
<tr>
<td>□ Television</td>
<td>Mr. &amp; Mrs. Smith / Alfred Hitchcock; Norman Krasna; Carole Lombard; Robert Montgomery; Gene Raymond; Jack Carson</td>
</tr>
<tr>
<td></td>
<td>Much ado about nothing / William Shakespeare</td>
</tr>
<tr>
<td></td>
<td>An old-fashioned girl / Louisa May Alcott</td>
</tr>
<tr>
<td></td>
<td>Pride and prejudice / Jane Austen</td>
</tr>
<tr>
<td></td>
<td>Sense and sensibility / Jane Austen</td>
</tr>
<tr>
<td></td>
<td>La tía Julia y el escribidor / Mario Vargas Llosa</td>
</tr>
</tbody>
</table>

Figure 1. Modified DeweyBrowser interface

The results include novels, films, and television programs. When a Dewey Category is selected, for example, Movies and films, the results are limited to items in that category:

High society
Mr. & Mrs. Smith

When a title is selected the work record for that title is displayed. The prototype will be publicly accessible and users will be invited to provide feedback on the interface and functionality.
5. Conclusion

Experience with FictionFinder and DeweyBrowser has shown that no single approach is adequate for finding and representing works with imaginary content. Changes in cataloging rules, formats, and practices have resulted in inconsistencies in the ways fictional prose and mixed formats are described and coded. The FictionFinder prototype demonstrates that, although a narrow definition of fiction which relies on a limited set of codes in MARC records can produce useful results, works of the imagination can be more fully represented when supplemented by content classed in the 700s and 800s as in the DeweyBrowser.

Notes
1 DDC, Dewey, and Dewey Decimal Classification are registered trademarks of OCLC Online Computer Library Center, Inc.
2 A new interface under development uses the term “edition” instead of the term “version” which was used in the original FictionFinder interface. For sample screens showing the new interface, see http://www.oclc.org/research/presentations/vizine-goetz/webwise2006.ppt.
3 Prior to the definition of 11 new codes and the change of name from “Fiction” to “Literary form” in 1997, only the codes 0 (Not fiction) and 1 (Fiction) were used. Retrieved February 28, 2006, http://www.loc.gov/marc/marbi/1996/96-8rrp2.html

References


